Case C→Q Checkpoint

The first two questions refer to the following information:

A Canadian study measuring depression level in teens (as reported in the *Journal of Adolescence*, vol. 25, 2002) randomly sampled 112 male teens and 101 female teens, and scored them on a common depression scale (higher score representing more depression). The researchers suspected that the mean depression score for male teens is higher than for female teens, and wanted to check whether data would support this hypothesis.

Question (1)

If μ_1 and μ_2 represent the mean depression score for male teens and female teens respectively, which of the following is the appropriate pair of hypotheses in this case?

A: $H_0: \mu_1 - \mu_2 = 0$
$H_a: \mu_1 - \mu_2 < 0$

B: $H_0: \mu_1 - \mu_2 > 0$
$H_a: \mu_1 - \mu_2 = 0$

C: $H_0: \mu_1 = \mu_2$
$H_a: \mu_1 > \mu_2$

D: $H_0: \mu_1 - \mu_2 = 0$
$H_a: \mu_1 - \mu_2 > 0$

E: Both (C) and (D) are correct.

Feedback

A : 0

✗ This is not quite right. Remember that you want to see if the mean depression score for male teens is ***higher*** than for female teens. Consider the remaining options. (E) is the correct answer.

B : 0

✗ This is not quite right. In general, the null hypothesis deals with equality among quantitative variables and the alternative hypothesis deals with inequality. Consider the remaining options. (E) is the correct answer.

C : 0

✗ This is not quite right. Although this option does work in this case, there is a better option. Consider the remaining options. (E) is the correct answer.

D : 0

✗ This is not quite right. Although this option does work in this case, there is a better option. Consider the remaining options. (E) is the correct answer.

E : 10

✓ Good job! We are comparing the mean depression scores of males (μ_1) and females (μ_2). Since we're hypothesizing that the males' mean is higher, the appropriate hypotheses are: H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 > \mu_2$ (as in (C)), which is equivalent to: H_0: $\mu_1 - \mu_2 = 0$ vs. H_1: $\mu_1 - \mu_2 > 0$ (as in (D)).

Question (2)

The following is the (edited) output for the test:

Two-Sample T-Test and CI

```
Sample     N   Mean   StDev   SE Mean
1(M)     112   7.38    6.95      0.66
2(F)     101   7.15    6.31      0.63


Difference = mu (1) - mu (2)
Estimate for difference:  0.230000
95% lower bound for difference:  -1.271079
T-Test of difference          :  T-Value = 0.25  P-Value = 0.400  DF = 210
```

From the output we learn that:

A: the data provide sufficient evidence to reject H_0 and to conclude that the mean depression score for male teens is larger than that of female teens.

B: the data provide sufficient evidence to conclude that male and female teens do not differ in mean depression score.

C: the data do not provide sufficient evidence to conclude that the mean depression score of male teens is larger than that of female teens.

D: the data do not provide sufficient evidence to reject H_0, so we accept it, and conclude that male and female teens do not differ in mean depression score.

Feedback

A : 0

✗ This is not quite right. Notice that the p-value is much larger than .05. Consider the remaining options. (C) is the correct answer.

B : 0

✗ This is not quite right. Since the p-value is large (0.4), the data do not provide enough evidence to reject the null hypothesis. Consider the remaining options. (C) is the correct answer.

C : 10

✓ Good job! Since the p-value is large (0.4) the data do not provide enough evidence to reject the null hypothesis in favor of the alternative hypothesis (which claims that male teens' mean depression score is higher). Note that (D) is not correct since â€œnot having enough evidence to reject H_0â€ does not mean that we accept it. In fact, as the course stresses many times, in the methodology of hypothesis testing, we *never accept* H_0.

D : 0

✗ This is not quite right. We never accept the null hypothesis. Consider the remaining options. (C) is the correct answer.

Grain is fortified with vitamins at the factory when processed. But, before the product reaches the consumer, some of the vitamins may degrade due to time, heat during storage, etc. Suppose the vitamin contents (in milligrams per pound) of five bags of grain are measured at the factory before shipping, and then again at the retail store after shipping. The results are as shown:

Bag	Vitamin content before shipping	Vitamin content after shipping
1	45	38
2	47	45
3	48	48
4	38	35
5	48	39

We wish to test whether there is a statistically significant decrease in vitamin content after shipping.

Question (3)

Given the design of the study and the question of interest, which of the following 4 computer outputs is relevant to use?

A:

Paired T-Test and CI: before shipping, after shipping

```
Paired T for before shipping - after shipping

                 N     Mean    StDev  SE Mean
before shipping  5  45.2000   4.2071   1.8815
after shipping   5  41.0000   5.3385   2.3875
Difference       5  4.20000  3.70135  1.65529


95% lower bound for mean difference: 0.67117
T-Test of mean difference = 0 (vs > 0): T-Value = 2.54  P-Value = 0.032
```

B:

Two-Sample T-Test and CI: before shipping, after shipping

```
Two-sample T for before shipping vs after shipping

                 N   Mean  StDev  SE Mean
before shipping  5  45.20   4.21      1.9
after shipping   5  41.00   5.34      2.4


Difference = mu (before shipping) - mu (after shipping)
Estimate for difference:  4.20000
95% lower bound for difference:  -1.55902
T-Test of difference = 0 (vs >): T-Value = 1.38  P-Value = 0.105
```

C:

Paired T-Test and CI: before shipping, after shipping

```
Paired T for before shipping - after shipping

                  N     Mean    StDev  SE Mean
before shipping   5  45.2000   4.2071   1.8815
after shipping    5  41.0000   5.3385   2.3875
Difference        5  4.20000  3.70135  1.65529


95% upper bound for mean difference: 7.72883
T-Test of mean difference = 0 (vs < 0): T-Value = 2.54  P-Value = 0.968
```

D:

Two-Sample T-Test and CI: before shipping, after shipping

```
Two-sample T for before shipping vs after shipping

                  N   Mean  StDev  SE Mean
before shipping   5  45.20   4.21      1.9
after shipping    5  41.00   5.34      2.4


Difference = mu (before shipping) - mu (after shipping)
Estimate for difference:  4.20000
95% upper bound for difference:  9.95902
T-Test of difference = 0 (vs <): T-Value = 1.38  P-Value = 0.895
```

Feedback

A : 10

✓ Good job! Since each of the 5 bags was measured twice (before and then after shipping), the appropriate inferential method for analyzing the data is matched pairs (answers A and C). The output in A is correct, since it tests the correct alternative. Since we would like to test whether the vitamin content decreases by the time the product reaches the consumer (i.e., before > after) the correct alternative is that the mean of the differences is > 0 ($\mu_d > 0$).

B : 0

✗ This is not quite right. Notice that the data are paired in the sense that the same variable was measured twice on each subject, before and after a treatment. You want to see if the final measurement is less than the initial measurement. Consider the remaining options. (A) is the right answer.

C : 0

✗ That is not quite right. Notice that the alternative hypothesis should involve the initial vitamin content being larger than the final vitamin content. Consider the remaining options. (A) is the right answer.

D : 0

✗ That is not quite right. Notice that the data are paired in the sense that the same variable was measured twice on each subject, before and after a treatment. You want to see if the final measurement is less than the initial measurement. Consider the remaining options. (A) is the right answer.

The next three questions refer to the following information:

To determine the relative effectiveness of different study strategies for the SAT, suppose three groups of students are randomly selected: One group took the SAT without any prior studying; the second group took the SAT after studying on their own from a common study booklet available in the bookstore; and the third group took the SAT after completing a paid summer study session from a private test-prep company. The means and standard deviations of the resulting SAT scores from this hypothetical study are summarized below:

	n	$\bar{x}$	s
Group 1 (no study)	12	1014.1	4.9
Group 2 (personal study)	12	1015.8	5.1
Group 3 (paid preparation)	9	1023.7	5.7

Since we are comparing more than 2 groups, we will use ANOVA to test whether the data provide evidence that SAT score is related to study strategy.

Question (4)

If we let μ_1, μ_2, and μ_3 be the mean SAT scores for students who use learning strategies 1, 2, and 3, respectively, the appropriate hypotheses in this case are:

A:
$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_a: \mu_1 \neq \mu_2 \neq \mu_3$$

B:
$$H_0: \mu_1 \neq \mu_2 \neq \mu_3$$
$$H_a: \mu_1 = \mu_2 = \mu_3$$

C:

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_a: \mu_1, \mu_2, \mu_3$ *are not all equal*

D:

$H_0: \mu_1, \mu_2, \mu_3$ *are not all equal*

$H_a: \mu_1 = \mu_2 = \mu_3$

E: Both (A) and (C) are correct.

F: Both (B) and (D) are correct.

Feedback

A : 0

✗ This is not quite right. Remember that you want to know if there is sufficient evidence to conclude that not all three groups scored the same on the SAT. However, this would still allow two groups to score equally well, while a third group did better or worse. Consider the remaining options. (C) is the right answer.

B : 0

✗ For ANOVA, the null hypothesis is that all of the groups' means are equal. Consider the remaining options. (C) is the right answer.

C : 10

✓ Good job! Note, as the course stresses, that (A) and (C) are not the same (and therefore option (E) is not correct).

D : 0

✗ That is not quite right. Remember that the null hypothesis for ANOVA is that all of the groups' means are equal. Consider the remaining options. (C) is the right answer.

E : 0

✗ That is not quite right. Option (A) and option (C) are not saying the same thing. Option (A) has the alternative hypothesis saying that all of the means are not equal. Option (C) has the alternative hypothesis saying that not all of the means are equal (in other words, some of the means can be equal). Consider the remaining options. (C) is the right answer.

F : 0

✗ That is not quite right. Remember the null hypothesis for ANOVA is that all of the groups' means are equal. Consider the remaining options. (C) is the right answer.

Question (5)

One of the conditions that allows us to use ANOVA safely is that of equal (population) standard deviations. Can we assume that this condition is met in this case?

A: No, since the three sample standard deviations are not all equal.

B: No, since the population standard deviations are not given, so we cannot check this condition.

C: Yes, since 5.7 - 4.9 < 2.

D: Yes, since 5.7 / 4.9 < 2.

Feedback

A : 0

This is not quite right. We consider the three sample standard deviations to be equal if the ratio of the largest sample standard deviation to the smallest sample standard deviation is less than 2. Consider the remaining options. (D) is the right answer.

B : 0

This is not quite right. Although the population standard deviations are not given, we can approximate them using the sample standard deviations. Consider the remaining options. (D) is the right answer.

C : 0

This is not quite right. We consider the three sample standard deviations to be equal if the ratio of the largest sample standard deviation to the smallest sample standard deviation is less than 2. Consider the remaining options. (D) is the right answer.

D : 10

Good job! The rule of thumb for meeting the condition of equal standard deviations is: largest SD / smallest SD < 2. In this case, we can assume that the condition is met, since the largest sample SD is 5.7, the smallest SD is 4.9 and 5.7 / 4.9 < 2.

Question (6)

Using the following output:

Analysis of Variance for SAT

	DF	SS	MS	F	P
Group	2	625.2	312.6	11.43	0.000
Error	30	820.7	27.4		
Total	32	1445.9			

we can conclude that:

A: the data provide strong evidence that SAT scores are related to learning strategy.

B: the data provide strong evidence that SAT scores are related to learning strategy in the following way: The mean SAT score for students who pay for coaching is higher than the mean SAT score for students who study themselves, which in turn is higher than that of students who do not study for the test.

C: the data provide strong evidence that the three mean SAT scores (representing the three learning strategies) are not all equal.

D: the data do not provide sufficient evidence that SAT scores are related to learning strategy.

E: Both (A) and (C) are correct.

Feedback

A : 0

That is not quite right. Although this statement is valid, there is a better option. Consider the remaining options. (E) is the right answer.

B : 0

That is not quite right. Although ANOVA can tell you whether or not all of the means are equal, it cannot tell you how they are ordered (it cannot tell you which way of preparing for the SAT works best). Consider the remaining options. (E) is the right answer.

D : 0

That is not quite right. The p-value is extremely small (approximately 0). As a result, we find that the data do provide sufficient evidence to suggest that SAT scores are related to learning strategy. Consider the remaining options. (E) is the right answer.

C : 0

That is not quite right. Although this statement is valid, there is a better option. Consider the remaining options. (E) is the right answer.

E : 10

Good job! Both statements (A) and (C) are correct. Since the p-value is essentially 0, we have strong evidence to reject H_0 and in favor of the alternative. The alternative hypothesis claims that the three SAT means (of the three learning strategies) are not all equal. In other words, we have strong evidence that performance on the SATs is related to learning strategy.

Question (7)

A GRE course prep company advertises that its course can significantly raise students GRE Verbal Reasoning scores. In order to assess the accuracy of this claim, a researcher randomly selected a sample of 30 graduating seniors from a local university, who are applying to business school. The 30 students first took the GRE, without any preparation. The 30 students then took the GRE again, after completing the GRE prep course. The researcher then compared the performance for each student to determine whether there was improvement on the GRE Verbal Reasoning scores from the first to second administrations.

What hypothesis testing technique should the researcher use to analyze the data?

A: ANOVA
B: z-test for the population mean
C: Two-sample t-test
D: Paired t-test

Feedback

A : 0

✗ That's incorrect. First, ANOVA is used when the categorical explanatory variable has more than 2 categories (i.e., when we're comparing more than two groups) and here we are comparing only two. Also, in ANOVA the samples are independent which is not the case here. Since the same group of students was measured twice (before and then after the GRE prep course) the samples are not independent. Therefore the appropriate statistical technique would be the paired t-test and the correct answer is D.

B : 0

✗ That's incorrect. The z-test for the population mean is used when we have only one population and you're conducting a test about the value for its mean (assuming is known). Here you're comparing two groups (those who take the GRE without the prep course and those who take the GRE after taking the prep course) using samples that are matched (since each students is measured twice; before and then after taking the prep course). Therefore, the appropriate statistical technique would be the paired t-test. The correct answer is D.

C : 0

✗ That's incorrect. While it is correct that we're comparing two groups, the samples are not independent since the same group of students was measured twice (before and then after the GRE prep course). Therefore, the appropriate statistical technique would be the paired t-test. The correct answer is D.

D : 10

✓ Correct! Since the study uses the same participants for both a pretest measure (before the GRE prep course) and a posttest

✓ (after the prep course) measure, the appropriate statistical technique would be the paired t-test.

Question (8)

In a hypothetical study, a pharmaceutical company is interested in finding the dosage of a new medication that best treats panic attacks. Ninety people, who are diagnosed with panic attacks, were randomly assigned to three groups: 1) 20 milligrams of the new drug, 2) 30 milligrams of the new drug, and 3) 40 milligrams of the new drug. After 30 days on the medication, each participant records the number of panic attacks they have experienced in the previous seven days.

What hypothesis testing technique should the researcher use to analyze the data?

A: ANOVA
B: z-test for the population mean
C: Two-sample t-test
D: Paired t-test

Feedback

A : 10

✓ Correct! We're comparing more than two (three) dosage groups, and since the subjects were randomized to one of the three dosage treatments, the three samples are independent. Therefore, it would be appropriate to use the ANOVA.

B : 0

✗ That's incorrect. The z-test for the population mean is used when we have only one population and you're conducting a test about the value for its mean (assuming is known). Here, you are comparing three dosage groups with respect to a quantitative variable (number of panic attacks) based on data collected from three independent samples. Therefore, the

✗ appropriate inferential method is ANOVA and the correct answer is A.

C : 0

✗ That's incorrect. The two-sample t-test is used is used when we're comparing two population means using independent samples. Here, we're comparing *three* dosage groups with respect based on data collected from three independent samples. Therefore, the appropriate inferential method is ANOVA and the correct answer is A.

D : 0

✗ That's incorrect. The paired t-test is used, when we comparing two population means using matched (non-independent) samples. Here, we're comparing *three* dosage groups with respect based on data collected from three *independent* samples. Therefore, the appropriate inferential method is ANOVA and the correct answer is A.

Question (9)

Do men and women differ, on average, in terms of the amount of television that they watch each day? A researcher conducted a hypothetical study, where he randomly selected 50 men and 50 women and recorded the number of minutes of television watched during the previous day.

Which of the following should the researcher use to determine whether there is a difference in mean number of minutes of television viewing between men and women?

A: z-test for the population mean
B: Two-sample t-test
C: Paired t-test
D: ANOVA

Feedback

A : 0

✗ That is incorrect. The z-test for the population mean is used when we have only one population and you're conducting a test about the value for its mean (assuming is known). In this case we're comparing two populations (males, females) with respect to the mean TV watching time based on data collected from two independent samples. The appropriate inferential method is therefore the two-sample t-test, and the correct answer is B.

B : 10

✓ Correct! We're comparing the two gender populations with respect to the mean TV watching time based on data collected from independent samples. The appropriate inferential method is therefore the two-sample t-test.

C : 0

✗ That's incorrect. The paired t-test is used, when the observations in one sample are linked to observations in another sample. Since the study involves random samples, the appropriate technique would be the two-sample t-test and the correct answer is B.

D : 0

✗ That's incorrect. ANOVA is used when we're comparing more than two population means based on data collected from independent samples. Note that in this study we're comparing the *two* gender populations' mean TV watching time based on data collected from independent samples. The appropriate analysis in this case is therefore the two-sample t-test, and the correct answer is B.